Assessment of the swine protein-annotated oligonucleotide microarray

J. P. Steibel*,[†], **M. Wysocki**[‡], **J. K. Lunney**[‡], **A. M. Ramos***,[1], **Z.-L. Hu**[§], **M. F. Rothschild**[§] and **C. W. Ernst***

*Department of Animal Science, Michigan State University, East Lansing, MI 48824, USA. [†]Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824, USA. [‡]Animal Parasitic Diseases Laboratory, ANRI, BARC, ARS, USDA, Beltsville, MD 20705, USA. [§]Department of Animal Science, Center for Integrated Animal Genomics, Iowa State University, Ames, IA 50011, USA

Summary

The specificity and utility of the swine protein-annotated oligonucleotide microarray, or Pigoligoarray (http://www.pigoligoarray.org), has been evaluated by profiling the expression of transcripts from four porcine tissues. Tools for comparative analyses of expression on the Pigoligoarray were developed including HGNC identities and comparative mapping alignments with human orthologs. Hybridization results based on the Pigoligoarray's sets of control, perfect match (PM) and deliberate mismatch (MM) probes provide an important means of assessing non-specific hybridization. Simple descriptive diagnostic analyses of PM/MM probe sets are introduced in this paper as useful tools for detecting non-specific hybridization. Samples of RNA from liver, brain stem, longissimus dorsi muscle and uterine endothelium from four pigs were prepared and hybridized to the arrays. Of the total 20 400 oligonucleotides on the Pigoligoarray, 12 429 transcripts were putatively differentially expressed (DE). Analyses for tissue-specific expression [over-expressed in one tissue with respect to all the remaining three tissues ($q < 0.01$)] identified 958 DE transcripts in liver, 726 in muscle, 286 in uterine endothelium and 1027 in brain stem. These hybridization results were confirmed by quantitative PCR (QPCR) expression patterns for a subset of genes after affirming that cDNA and amplified antisense RNA (aRNA) exhibited similar QPCR results. Comparison to human ortholog expression confirmed the value of this array for experiments of both agricultural importance and for tests using pigs as a biomedical model for human disease.

Keywords amplified antisense RNA, long oligo array, pig, quantitative PCR.

Introduction

DNA microarrays allow the simultaneous evaluation of transcriptional profiles for thousands of genes. Whole genome DNA microarrays in particular are used to assess the effect of multi or single factorial perturbations on the transcriptome of one or more types of cells. A popular implementation of whole genome arrays are long oligonucleotide microarrays. These arrays are composed of 40- to 70-mer oligonucleotides spotted on a glass slide. Desirable properties of these arrays are efficient hybridization to the target probe and simultaneous low cross-hybridization (Zhao *et al.* 2005; Tuggle *et al.* 2007). Other desirable properties of a whole genome array are comprehensive coverage of the transcriptome over a range of tissues and conditions, and availability of annotation information.

As the pig is both an important agricultural species and a good comparative model for biomedical research, a number of whole genome microarray resources have been generated (Zhao *et al.* 2005; Tsai *et al.* 2006; Lunney 2007; Wang *et al.* 2007). More recently, an improved long oligonucleotide microarray has been released to the research community as a result of collaborative efforts among pig and cattle genome researchers. The new 70-mer oligonucleotide microarray is comprised of 20 400 oligos; the Swine

Address for correspondence

J. P. Steibel, Department of Animal Science, Michigan State University, East Lansing, MI 48824, USA.
E-mail: steibelj@msu.edu

[1]Present address: Animal Breeding and Genomics Centre, Wageningen University, Wageningen, the Netherlands.

Protein-Annotated Oligonucleotide Microarray, or Pigoligoarray, has been designed based on known swine expressed sequences and annotated using human, cattle and porcine protein information (http://www.pigoligoarray.org). The Pigoligoarray also includes a set of control probes to assess non-specific hybridization.

In this study, we evaluate the Pigoligoarray by profiling the expression of transcripts in four porcine tissues. We validate the hybridization results by comparative analysis of expression in human orthologs, confirm expression patterns for a subset of genes by quantitative PCR (QPCR), and assess the usefulness of designed control oligonucleotides. Finally, we assess the sensitivity of the microarray to detect tissue enhanced gene expression, comparing results to ortholog expression in human tissues and related species.

Materials and methods

Oligonucleotide microarray

This study was conducted to evaluate the newly developed Swine Protein-Annotated Oligonucleotide Microarray, or Pigoligoarray, a second generation porcine 70-mer oligonucleotide array. This microarray was developed as an open source collaboration between investigators and institutions interested in pig physiology (Swine NRSP-8/NC1037 Committee led by Dr. S. Fahrenkrug, University of Minnesota). The oligonucleotide probes developed by Illumina were designed from contigs developed by Dr. C. Elsik (Georgetown University) by comparison of pig expressed sequence tags (ESTs) to phylogenetically defined vertebrate proteins. Oligonucleotides were annotated using descriptions of homologous proteins; information regarding the oligonucleotides, consensus sequences and availability of printed arrays can be found at http://www.pigoligoarray.org. Briefly, the microarray includes 20 400 70-mer oligonucleotides. Hybridization stringency controls include six mismatch (MM) probes (1, 2, 3, 5, 7 and 10 MMs) designed against each of 60 contigs with the highest EST count in the database. There are 60 negative control oligonucleotides that correspond to scrambled sequences without presumed representation in either the pig or bovine genome or pig EST databases. As part of this research report, Pigoligoarray oligonucleotides were further annotated, noting their HUGO Gene Nomenclature Committee (HGNC) identities http://www.genenames.org, gene identity numbers and Gene Ontology (GO) categories http://www.geneontology.org, as well as their projected mapping alignments compared with their human orthologs.

Tissue sampling, RNA isolation and labelling

The main experiment utilized samples collected from four pigs (unrelated gilts at approximately 165 days of age) that were slaughtered in the federally inspected Michigan State University (MSU) Meats Laboratory. Samples of liver, brain stem, longissimus dorsi muscle and uterine endothelium were flash frozen in liquid nitrogen and stored at −80 °C. Animal experimental procedures were approved by the MSU Institutional Animal Care and Use Committee (11/04-141-00). Total RNA from 1.0 g of each tissue sample was extracted using TRIzol reagent (Invitrogen Corp.) according to the manufacturer's instructions. RNA concentration and quality was determined with an RNA 6000 Pico LabChip® kit using an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc.). Additional samples were used for the background hybridization and RNA/aRNA (amplified antisense RNA) comparison studies; these included foetal muscle tissues collected at MSU and palatine tonsils, tracheal bronchial lymph nodes (TBLN) and lung samples at BARC collected for related studies (Petry *et al.* 2007; Wysocki *et al.*, unpublished data). All tissues were stored at −80 °C and RNA was extracted and quantified as described above.

Tissue samples from the four gilts were evaluated at MSU in a connected loop design (Rosa *et al.* 2005) microarray experiment (Fig. S1). Four loops were used, one for each animal, such that loop and animal were confounded. However, the tissue sequence between loops was altered such that all tissue pairs were represented in at least one array, and tissue and dye levels were balanced. Results were compared with data from additional experiments (Table S1).

For each sample, 1 µg of total RNA was reverse transcribed with a T7 oligo(dT) primer using the Amino Allyl MessageAmp™ II aRNA Amplification Kit (Ambion Inc.) according to the manufacturer's instructions. Following first-strand and second-strand synthesis and purification, the cDNAs were *in vitro* transcribed to synthesize multiple copies of amino allyl-modified aRNAs. After aRNA purification, some aRNAs were used for the RNA/aRNA comparison; 10 µg of aRNAs were labelled with N-hydroxysuccinate (NHS) ester Cy3 or Cy5 dyes (GE Healthcare) as appropriate for the experimental design. The labelled aRNAs (2.5 µg) were purified and combined with 65 µl of Slide Hyb #1 solution (Ambion Inc.) and denatured at 70 °C for 5 min. Additional microarrays from other studies (Table S1) performed at MSU or at BARC were used for assessing usefulness of control features included on the arrays. At BARC, a similar labelling protocol was followed using Alexa Fluor® 555 and Alexa Fluor® 647 dyes (Invitrogen).

Microarray hybridization and image processing

Pigoligoarray hybridizations were performed in sealed hybridization cassettes (ArrayIt, TeleChem International, Inc.) for 18 h at a humid 54 °C. A low stringency experiment was also conducted at MSU with hybridizations at 42 °C. Following hybridization, slides were washed in 2× SSC/0.5% SDS and 0.1× SSC/0.1% SDS solutions for 10 min each. The slides were rinsed in a 0.1× SSC solution and

nuclease-free water and dried by centrifugation. Fluorescent images were detected by an Axon GenePix® 4000B scanner (Molecular Devices), and fluorescence intensity data were collected using GENEPIX® PRO 6 software (Molecular Devices) after spot alignment. The dataset was submitted to the National Center for Biotechnology Information's Gene Expression Omnibus database [GEO: GSE13095]. Median intensity values for each dye channel were stored as comma-separated values data files. Median intensities were extracted and normalized using a within- print-tip lowess location normalization and an overall scale normalization (Yang *et al.* 2002). This normalization removed intensity dependent biases from each printing block in each slide. The resulting normalized data were expressed in the $\log_2$ scale.

QPCR analysis

Eleven candidate genes were selected *a priori* (i.e. before knowing the results of the microarray analysis) for Pigoligoarray confirmation based on a combination of (i) their known expression in humans (based on GeneCard data for 12 tissues: http://www.genecards.org), (ii) their function, (iii) the availability of the pig genomic sequence and (iv) potential involvement in disorders and diseases in humans and animals. Based on previous data (Zhao *et al.* 2005), four genes (*MAPK1, INDO, IRF2, STAT6*) were selected as candidates, as they were analysed for the first long oligo array (Qiagen) in four pig tissues (liver, lungs, small intestine and muscle), and thus their expression could be compared with the new Pigoligoarray data. Three additional genes were selected for QPCR confirmation *a posteriori* from results of the microarray analysis.

Probes and primers were designed using the Primer Express (Applied Biosystems) software. All nucleotide sequences were obtained from NCBI Entrez Nucleotide database http://www.ncbi.nlm.nih.gov/sites/entrez?db=nuccore&itool=toolbar or the Pig Expression Data Explorer database (http://pede.dna.affrc.go.jp; Uenishi *et al.* 2004, 2007). Exon–exon junctions were identified by comparisons to the human genome sequence, so that genomic DNA contamination would not appear as additional PCR products and that primers and probes would only amplify unique sequence. Sequences for probes and primers are available at the Porcine Immunology and Nutrition database (http://www.ars.usda.gov/Services/docs.htm?docid= 6065; Dawson *et al.* 2005). Synthesis of cDNA was performed using Superscript reverse transcriptase (Invitrogen) and oligo dT with 5 μg of total RNA for the validation study or with 2 μg of total RNA for the RNA/aRNA comparison study; aRNA was transcribed with the Ambion kit as noted above. All samples were measured in duplicate. QPCR amplification reactions were carried out using the Brilliant kit (Stratagene) and ABI Prism 7500 sequence detector system (Applied Biosystems) as previously described (Dawson *et al.* 2005). The thermal cycling programme included two stages: 95 °C for 10 min, and 40 cycles of 95 °C for 15 s and 60 °C for 1 min.

Statistical analysis of differential gene expression

For differential expression analysis of the Pigoligoarray data, normalized (log) intensities were analysed using the following oligonucleotide-specific linear mixed model,

$$y_{ijkl} = \mu + A_i + D_j + T_k + DT_{jk} + B_l + TB_{kl} + e_{ijkl};$$

where fixed effects included were dye (*D*), tissue (*T*) and their interaction (*DT*), and random effects comprised array (*A*), animal (*B*) and the interaction of animal with tissue (*TB*). An overall *F*-test of differential expression across tissues was computed as well as a deviation test comparing the mean expression in a tissue to the overall expression. Finally, all pairwise comparisons between tissues were obtained. For each test, *P*-values were adjusted to *q*-values using the false discovery rate (FDR) procedure (Storey & Tibshirani 2003), estimating the proportion of null hypotheses using a mixture model approach (Gadbury *et al.* 2004).

Cycle to threshold (C_t) QPCR data was analysed using a linear mixed model as described below.

$$C_{t_{kl}} = \mu + T_k + B_l + e_{kl};$$

where *T* is the fixed effect of tissue and *B* is the random effect of animal. A negative difference in C_t indicates a positive fold-change. We expect the results from this analysis to have similar magnitude, with opposite signs for the equivalent comparisons using log intensities in the microarray analysis.

Results

Comparative GO annotation of oligonucleotide set

To facilitate gene expression and pathway analyses, more detailed gene annotations are helpful. Thus efforts were made to add the official HGNC gene name, gene identity number and symbol (short-form abbreviation) for every Pigoligoarray oligonucleotide as well as identify their relevant GO terms. (Data available at http://www.animal genome.org/cgi-bin/host/Lunney/oligoAnnotatn.) Based on these assignments for the Pigoligoarray, a total of 86 340 GO class IDs have been assigned and were mapped to GO_slim ancestor terms http://www.geneontology.org/ GO_slims/ using CATEGORIZER (Hu *et al.* 2008; Results available at http://www.animalgenome.org/jlunney/share/ oligoAnnot/GOslim.php). The genome representation for the different pig arrays was compared based on GO class IDs. Information on the Affymetrix array is available at http:// www.affymetrix.com/products/arrays/specific/porcine.affx; it contains 23 937 probe sets to interrogate 23 256 transcripts in pig, which represents 20 201 genes, for

which we were able to assign 14 624 GO terms. The earlier swine oligonucleotide array, the NRSP8-Qiagen array, had 12 500 probes (Zhao *et al.* 2005) for which 5853 GO terms were assigned.

The detailed GO term lists are provided in Table S2 and summarized in Table 1. The coverage of the three arrays for high frequency GO terms is compared in Fig. 1a. It is clear that, based on available annotation, substantially more oligonucleotides were identified for the Pigoligoarray than for the Affymetrix or Qiagen arrays. This represents both the larger number of genes represented on the Pigoligoarray as well as the more accurate annotation of the genes for the 2006 Pigoligoarray design vs. the 2004 Affymetrix and 2003 Qiagen designs. A rapid comparison of the arrays indicated that the Pigoligoarray had many more representative oligonucleotides in general (Table 1) and for each category, e.g. for GO:0003674 (molecular function), it had 35 017 probes whereas the Affymetrix array had 7332 probes and the Qiagen array only 1759 probes (Fig. 1a). These differences were even more dramatic when low frequency GO terms were analysed (Fig. 1b), e.g. for GO:0007165 (signal transduction), the Pigoligoarray had 3274 oligos whereas Affymetrix had 487 and Qiagen only 207 oligos. For immune-related events, e.g. for GO: 0006950 (response to stress), the representation was 882, 115 and 117 and for GO:0008219 (cell death), 717, 93 and 85, for the Pigoligoarray, Affymetrix and Qiagen arrays respectively. For some terms, the Pigoligoarray had even greater differences, e.g. for GO:0016032 (viral life cycle), 46, 4 and 5 or for GO:0016209 (antioxidant activity), with 46, 30 and 3, for the Pigoligoarray, Affymetrix and Qiagen arrays respectively. Thus, the new Pigoligoarray has a larger number of probes representing a broader range of cellular functions. As an additional tool for researchers, all of the Pigoligoarray probes have been annotated with their HGNC assignments and shown on the human genome map with a comparative swine map alignment at http://www. animalgenome.org/cgi-bin/QTLdb/SS/link_oligo2hs.

Figure 1 Comparison of coverage of genes and related Gene Ontology (GO) terms for Qiagen, Affymetrix and Pigoligoarray microarrays. (a) Coverage of array oligonucleotides for high frequency GO terms: GO:0003674 (molecular function); GO:0008150 (biological process); GO:0005575 (cellular component); GO:0008152 (metabolism); GO:0003824 (catalytic activity); GO:0005488 (binding); and GO:0005623 (cell). (b) Coverage of array oligonucleotides for low frequency GO terms: GO:0030528 (transcription regulator activity); GO:0030154 (cell differentiation); GO:0006950 (response to stress); GO:0008219 (cell death) and GO:0009605 (response to external stimulus).

Characterization of control oligonucleotide hybridizations

A unique feature of the Pigoligoarray is the presence of negative probes and perfect match/MM (PM/MM) sets of probes. These probes constitute useful indicators of hybridization quality. In particular, the intensity from negatives and PM/MM sets can be used as indicators of overall non-specific binding. This is clear when relative signal intensity for negative and non-control oligos for 'good' experiments with high stringency hybridization conditions are compared to 'poor' ones with high non-specific hybridization (Fig. 2a compared to Fig. 2b). Negative control oligos have median signal intensities (A-value) similar to or lower than non-control oligos for the 'good' arrays (Fig. 2a); whereas for the 'poor' arrays, negatives exhibited an almost symmetric distribution of intensities around the median value, with the median of negatives in general below the median of non-control oligos in each array (Fig. 2b).

Table 1 Comparison of the number of GO annotations among three arrays in terms of gene coverage[1].

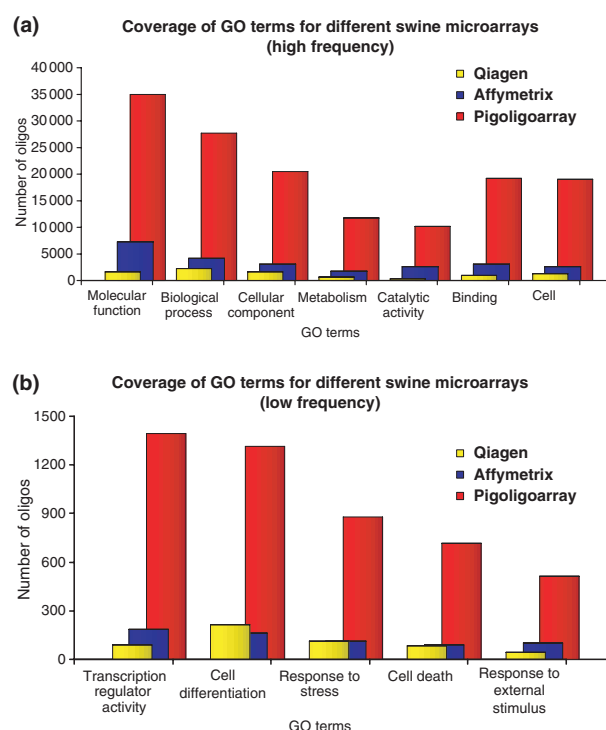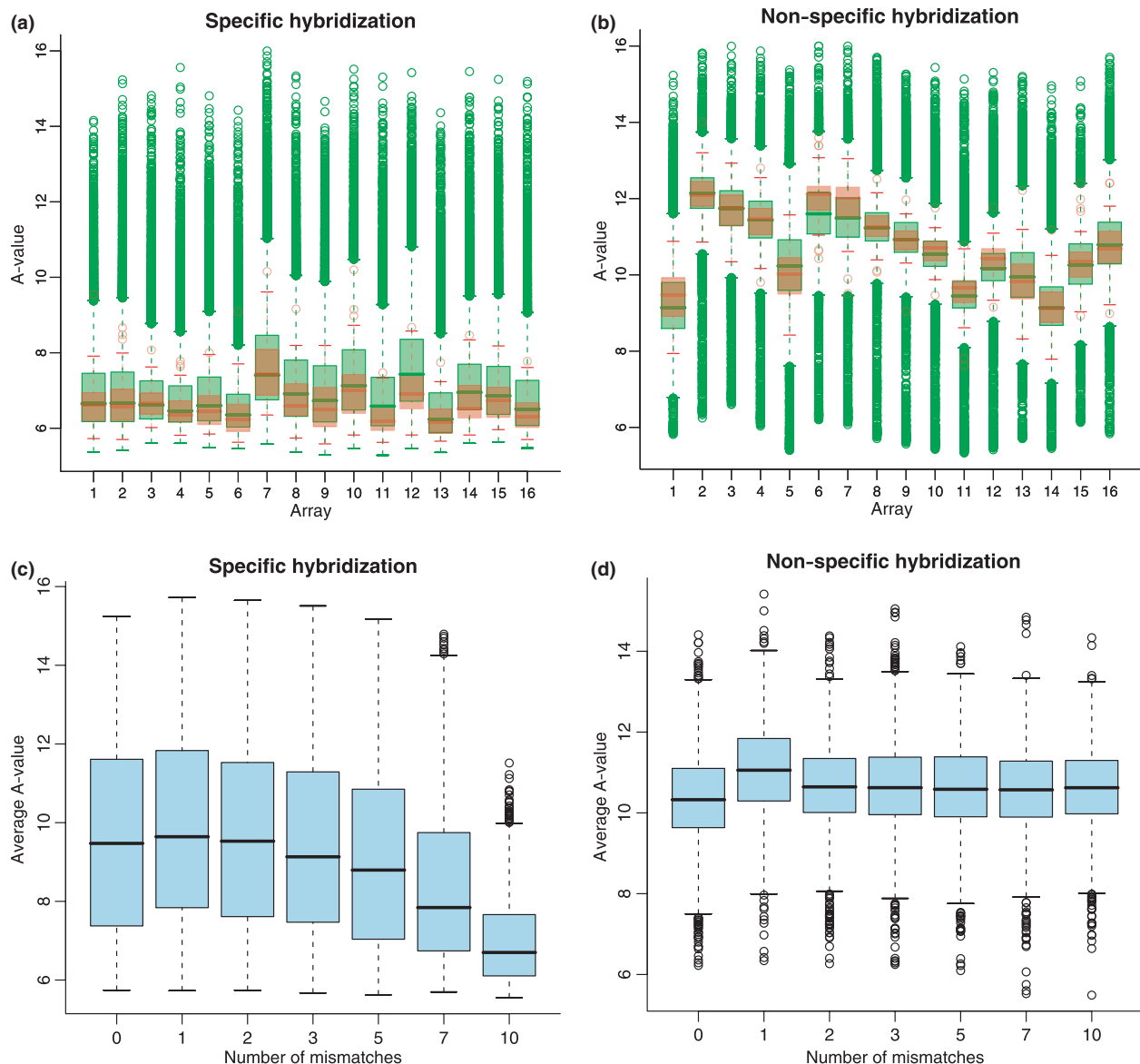| Oligo-set	Targets	GO annotated	Total GO IDs[2]	Unique GO IDs
Qiagen	13298	1265[3]	5853	1956
Affymetrix	24123	2440[4]	14624	1474
Pigoligoarray	18524	16225[5]	86340	4624

[1]Data on all GO terms are listed in Table S2.
[2]Multiple GO annotations exist per target.
[3]Qiagen GO annotation: by blast at e-3.
[4]Affymetrix probes GO annotation: 'Affy-annot.txt' downloaded from: http://www.affymetrix.com/products/arrays/specific/porcine.affx.
[5]Pigoligoarray oligo annotation: by the consortium: http://primer.ansci.umn.edu/pigoligoarray/annotation.htm.

Figure 2 Effect of stringency on hybridization intensity to negative (green boxes) and non-control oligonucleotides (red boxes, panels a and b) or to deliberate mismatched oligonucleotides (panels c and d). (a) Hybridization intensity (A-values) of negatives for experiment on 16 different arrays conducted with high stringency hybridization conditions; (b) experiment with widespread non-specific hybridization. (c) Overall average hybridization intensity of mismatched oligonucleotides (across arrays and oligonucleotides) in an experiment conducted under high stringency hybridization conditions; (d) experiment with widespread non-specific hybridization.

A problem arose when we first reviewed the results for the 60 negative oligonucleotides: we found consistently high intensity for a subset of six oligos. The results were not only consistent in the arrays from our assessment experiment, but also for hybridization intensities in another seven experiments involving a variety of tissues and experimental perturbations (Table S1). As these negatives were designed before substantial swine genome sequence was available, it was possible that these oligos represented previously unidentified swine sequences. However, a blast search did not reveal homology to known swine, human or mouse

gene sequences. Consequently, we cannot assert that these oligonucleotides are not true negatives based on their sequences, but the consistently high signals in numerous experiments provide evidence that they do not act as true negatives; therefore, these six oligonucleotides were excluded from the negative set for all downstream analyses (Fig. S2).

Across all PM/MM sets for the 'good' experiments, there was a decay in median hybridization signal as the number of MMs increased (Fig. 2c). Additionally, the dispersion around the median intensity decreased as the number of

MMs increased. These patterns were not present in arrays with non-specific hybridization (Fig. 2d).

In experiments with specific hybridization, individual probe analysis of MM oligo sets revealed the expected decrease in signal intensity (A-value) with increased MM number for the highly expressed probes (Fig. 3a). However for some probe sets this was not the case; these corresponded to six probes with very low signal, and thus the increase in MM numbers had limited effect (Fig. 3b).

(a) **Specific hybridization, expressed PM**

(b) **Specific hybridization, non-expressed PM**
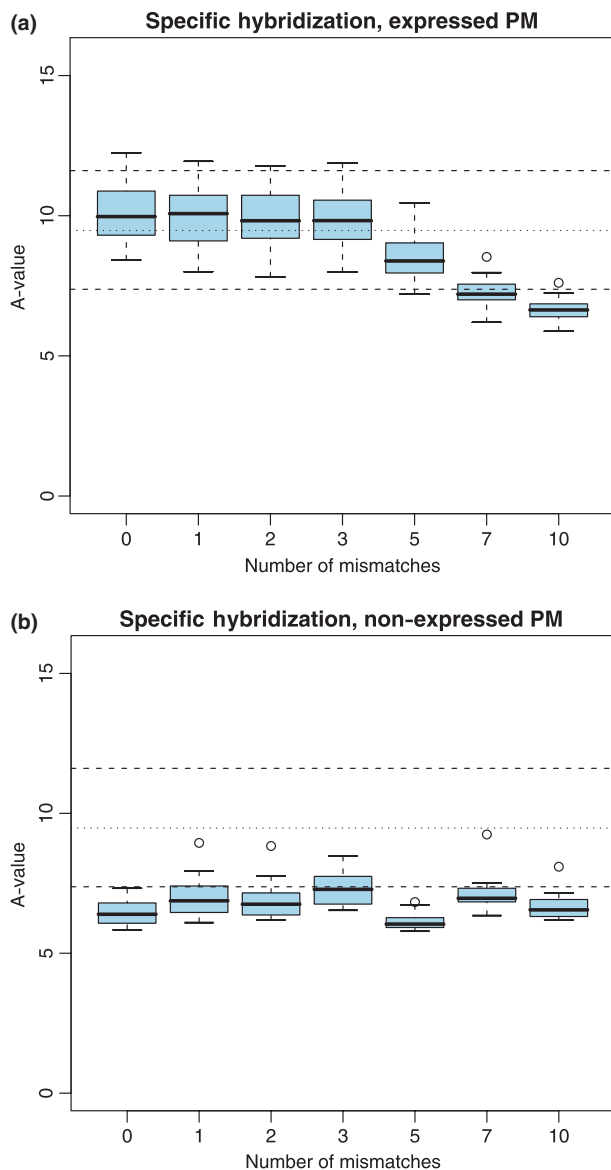
Figure 3 Effect of number of mismatches on average log-intensity. (a) The average hybridization intensity (across arrays) for perfect match (PM) oligos that are highly expressed; (b) for individual PM oligos that are not expressed in the target tissue. Horizontal lines represent the median (solid line) and quartiles (dashed lines) of the whole PM/MM set.

Transcriptome analysis

Samples of RNA from liver, brain stem, longissimus dorsi muscle and uterine endothelium from four pigs each were prepared and hybridized to arrays using a loop design (Fig. S1). Using a very stringent Bonferroni correction, 49 and 156 transcripts were detected significantly differentially expressed (DE) at $P < 0.01$ and $P < 0.05$ respectively. The low number of significant differences is expected considering the stringency of the Bonferroni correction, for example to call a significant difference at $P < 0.01$ after the correction; the nominal P-value has to be smaller than 4.82×10^{-7} (0.01/20736). The distribution of P-values, however, showed very strong evidence of tissue differential expression (Fig. S3). A more reasonable criterion to call DE transcripts in this case is the FDR (Storey & Tibshirani 2003). We estimated the proportion of null hypotheses as $\pi_O = 0.1064$, and computed the q-value (FDR equivalent to P-values) to find 12 429 transcripts putatively DE ($q < 0.01$). We expect only 124 of these to be false positives (Fig. S3b).

For each transcript, the expression in each tissue was compared against the expression in the other three tissues individually. We called a transcript tissue-specific if it was significantly over-expressed in one tissue with respect to all the remaining three tissues ($q < 0.01$). This yielded 958 transcripts in liver, 726 in muscle, 286 in uterine endothelium and 1027 in brain stem.

We ranked transcripts from the four tissue-specific lists by decreasing fold-change of the deviation contrast, selected the top 15 annotated transcripts in liver and muscle (largest positive fold-change per tissue relative to average of all tissues) and compared the expression profiles to the profiles obtained with Affymetrix GeneChips in equivalent human tissues (Su *et al.* 2004). We confirmed most of the expression differences in liver and skeletal muscle tissue for these transcripts. Comparisons in uterine endothelium and brain stem were not straightforward because the database did not include exactly the same type of tissue as our experiment. In total, out of 30 comparisons for muscle and liver, we observed agreement in 26 of them (Table S3).

Understanding the relative magnitude of variance components in two-colour microarray experiments is key to optimally design future transcriptional profiling experiments (Cui & Churchill 2003). We studied the distribution of the four variance components fitted in the mixed model. In general, the array variance accounted for most of the total variance (on average 76%), followed by the residual variance (9.5%). The animal and animal by tissue component explained a smaller proportion of the total variance compared to the rest of the variances (5.9% and 8.5% respectively). Moreover, the two 'biological variances' (corresponding to animal and animal by tissue effects) were estimated to be close to zero in a large proportion of transcripts (Fig. 4).
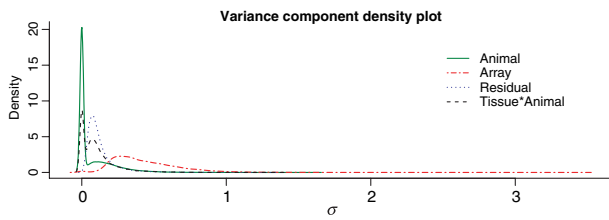
Figure 4 Distribution of gene-specific sources of variation. The distribution of four variance components is indicated to have very different relative magnitude of the variance associated with each experimental factor. The x-axis is the standard deviation and the y-axis represents the density (proportional to frequency) for a given level of x. Animal and Tissue*Animal components show spikes at zero. Array component shows a much higher value than the other components.

Comparative QPCR analysis

As QPCR assays were performed with cDNA, we first wanted to confirm that results with cDNA prepared from total RNA would be comparable to those from aRNA. Results for tests of 27 samples of both aRNA and RNA from independent sets of tissues (foetal muscle, palatine tonsils, TBLN and lung) are shown in Fig. 5a. Statistical analyses of the data show that, when equivalent amounts of starting RNA/aRNA are compared, there are no significant differences in gene expression for the genes analysed by QPCR using RNA compared with aRNA for nine of the eleven genes tested; we found significant differences between aRNA compared with RNA based C_t measures for the genes *IGF2* and *NFKBIA* ($P < 0.01$).

A set of 11 genes selected *a priori* (i.e. before knowing the results from the microarray analysis) plus an additional three genes selected *a posteriori* were evaluated in the four tissues using QPCR. The Pigoligoarray probes representing the 14 genes varied between one and three oligonucleotides. Differences in gene expression between the four experimental tissues across all 14 genes are presented in Fig. 5b. As expected, among the *a priori* selected genes, *MBP* is exclusively expressed in brain while the other genes exhibit a wide variety of expression levels across investigated tissues, confirming the adequate selection of candidate genes for this study. The genes *INDO, IRF2, MAPK1* and *STAT6* were selected to allow comparison to the first generation pig oligonucleotide microarray (Zhao *et al.* 2005). However, no significant differences were observed among the 4 tissues examined using the Pigoligoarray in this study for *IRF2, MAPK1* or *STAT6*. Also, while *INDO* did exhibit some significant tissue differences in this study, liver and longissimus dorsi muscle were not significantly different, and these two tissues were the only common tissues for the two array studies. Thus, the four selected genes did not provide a useful comparison of the two array validation studies largely due to the different tissues used for each study.

Results of tissue comparisons for the Pigoligoarray gene expression data and QPCR were contrasted; genes identified
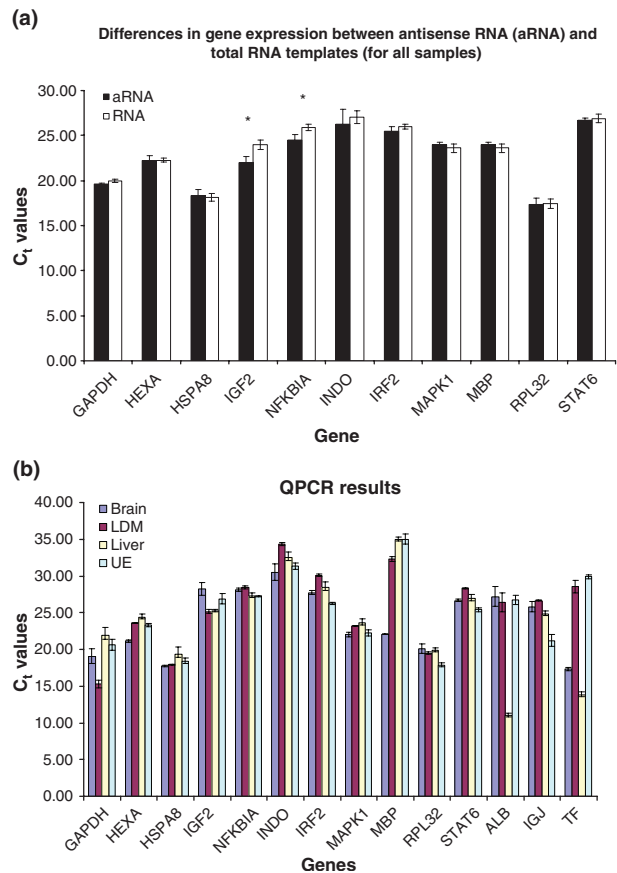

Figure 5 Comparison of quantitative PCR (QPCR) based gene expression. (a) Differences in gene expression between amplified antisense RNA (aRNA) and total RNA templates for 11 genes determined by quantitative PCR (QPCR) on 27 samples of skeletal muscle, tonsil, lymph node and lung. Bars represent mean ± SE. Genes with asterisk differ significantly ($P < 0.01$). (b) Expression of 14 genes in different tissues as measured by QPCR for 16 targeted gilt samples ($n = 4$ per tissue). LDM: longissimus dorsi muscle; UE: uterine endothelium. Bars represent average C_t values.

with the microarray to have significant differential expression ($P_{array} < 0.01$) are included in Table 2. An agreement between QPCR and array results was accepted when the overall pattern of differential expression detected by the microarray was confirmed by the QPCR assay. This usually involved the most extreme tissue expression as well as the overall correlation between both measurements. A negative correlation is expected between the C_t value and the log-intensity from the microarray. For those genes and comparisons that were indicated as significantly differentially expressed by the microarray experiment ($P < 0.01$, 10 genes), seven agreed in significance ($P < 0.01$) and overall pattern of differential expression with QPCR results. Three genes did not show significant correlation in their expression assayed by QPCR and microarray, but one of these (*NFKB1A*) tended to agree in at least two contrasts involving the relative expression in brain tissue (Table 2),

Table 2 Summary of QPCR and microarray results[1].

HGNC identity	QPCR results[2]					Microarray results[3]			Technical ρ[4]
	Brain	Liver	Muscle	UE	SEM	P-value	Significant diffs.	Confirmed	
ALB	26.9[a]	11.0[b]	26.4[a]	26.7[a]	1.36	<.0001	L High	Yes	−0.87**
GAPDH	18.7[b]	20.1[a]	14.4[c]	20.3[a]	0.33	0.0014	M High	Yes	−0.59**
HSPA8	17.8[b]	19.4[a]	18.0[b]	18.4[b]	0.44	0.003	B low	No	0.10
IGF-II	28.3[a]	25.2[b]	25.2[b]	26.9[ab]	0.79	0.0005	L High, B Low	Yes	−0.72**
IGJ	25.4[a]	24.9[a]	26.6[a]	21.2[b]	0.86	<.0001	UE High	Yes	−0.87**
INDO	30.5[b]	32.5[ba]	34.3[a]	31.3[b]	0.89	0.0005	L High	No	0.22
MBP	19.8[c]	34.3[a]	32.6[b]	34.5[a]	0.52	0.0002	B high	Yes	−0.61**
NFKBIA	28.1[a]	27.3[b]	28.4[a]	27.2[b]	0.26	0.0022	B < UE and L	Yes	−0.18
RPL32	19.6[a]	19.9[a]	19.3[a]	18.1[b]	0.25	0.002	UE high	Yes	−0.66**
TF	17.4[b]	13.8[b]	28.6[a]	29.9[a]	1.02	<.0001	L high	Yes	−0.80**

[1]RNA from four tissues, liver (L), longissimus muscle (M), uterine endothelium (UE) and brain stem (B), was compared for gene expression using the Pigoligoarray or by QPCR. Only genes found to be significantly differentially expressed ($P < 0.01$) with the Pigoligoarray are included in the table.
[2]C_t is cycles to threshold. Low C_t means high expression. Mean values sharing a letter do not differ ($P > 0.01$).
[3]If results of comparisons of gene expression for the array were in agreement to QPCR, then it is noted as Yes.
[4]Correlation between log-intensity and C_t value.
$\rho = -1$ indicates perfect correlation. **ρ is different from zero ($P < 0.001$).

and thus the pattern of *NFKB1A* expression was consistent between the microarray and QPCR. The discrepancies between significant differences found by the microarray and QPCR are more apparent for the genes *INDO* and *HSPA8*. Further inspection of the QPCR results revealed that *INDO* is expressed at a very low level (average $C_t = 32.16$) and consequently, we could expect the microarray results to be subject to more technical variation. The oligonucleotide on the Pigoligoarray showing sequence homology to the human *HSPA8* gene had in general low intensity compared with other transcripts; it appeared to be over-expressed in brain compared to uterine endothelium and longissimus dorsi muscle. These results contrasted with the QPCR results that indicated a highly expressed transcript with relatively consistent expression across the four tissues. We considered only significant differences for comparing microarray and QPCR results and did not filter the data by a minimum fold-change. For the two genes that were not confirmed by QPCR (*INDO* and *HSPA8*), relative fold-changes observed with the Pigoligoarray were low. Thus, if minimum fold-change criteria had been considered in addition to *P*-value, confirmation rate for significantly differentially expressed genes in this study would have been even higher.

Discussion

This report highlights the value of the new Pigoligoarray for analyses of swine gene expression differences. Data are presented on tissue-specific expression patterns as well as methods for assessing the quality of hybridization results. The oligonucleotides included on the Pigoligoarray have substantially better annotation compared with previous whole genome swine microarray resources as a result of the design, whereby a significant portion of the annotated

oligonucleotides correspond to known proteins. As part of this report, we have further improved the annotation of the Pigoligoarray, facilitating the comparison to human ortholog expression data available in the public domain.

Assessing the specificity of competitive hybridization is essential to obtain reliable microarray data. Control features included on the Pigoligoarray allow a rapid assessment of the overall hybridization specificity. In this paper, we show that a descriptive analysis of hybridization intensities of PM/MM sets was more informative than the analysis of hybridization results for negative oligonucleotides compared with non-controls. For negative oligos, median negative intensity may be very close to the median intensity of non-control oligos for some experiments, especially if a relatively small number of genes are expressed in the tissue being evaluated in the experiment. Additionally, some negative probes show consistently high intensity across experiments. On the other hand, PM/MM sets allow assessment of the decay in intensity as a function of the number of MMs within each probe set, such that each PM is a positive control for the corresponding set of MMs. If a particular PM probe is not expressed in one experiment, the decreasing intensity pattern will not be observed, and the whole PM/MM set can be safely discarded from the diagnostic analysis.

The microarray showed very good specificity for tissue selective gene expression. Based on a mixed model analysis that accounted for technical and biological sources of variation, we detected a large number of differentially expressed genes while controlling the FDR at a low level ($q = 0.01$). Using this microarray, we were able to detect several hundred tissue selective genes for each tissue, and confirmed such tissue selectivity in a subset of genes by comparing tissue

transcriptional profiles to human ortholog tissue-specific expression.

Genes selectively expressed in skeletal muscle included the major contractile protein genes *alpha actin* (*ACTC1*), *myosin heavy chain 1* (*MYH1*) and *myosin light chain 1* (*MYL1*) as well as genes encoding myofibrillar regulatory proteins such as *alpha tropomyosin* (*TPM1*) and cytoskeletal proteins such as *titin* (*TTN*). All of these transcripts are expected to be abundantly expressed in skeletal muscle based on the abundance of their protein products in myofibrils (Aberle *et al.* 2001). Other genes found to be selectively expressed in skeletal muscle are involved in calcium transport, including *sarcolipin* (*SLN*; Babu *et al.* 2007) and *triadin* (*TRDN*; Shen *et al.* 2007). In addition, several metabolic enzymes including *fructose-bisphosphate aldolase A* (*muscle-type aldolase, ALDOA*; Mukai *et al.* 1986), *muscle-specific carbonic anhydrase III* (*CA3*; Tweedie & Edwards 1989) and *adenosine monophosphate deaminase 1 (muscle isoform, AMPD1*; Morisaki & Holmes 1993), which are expected to be expressed at high levels in skeletal muscle, were observed to be selectively expressed in muscle.

Results for liver included many selectively expressed genes that were expected based on their functions and known tissue specificity in other species. These included several genes involved in the coagulation pathway such as *plasminogen* (*PLG*; Currier *et al.* 2003), *fibrinogen gamma chain* (*FGG*; Duan & Simpson-Haidaris 2006), *apolipoprotein H* (*APOH*; Ragusa *et al.* 2006) and *serpin peptidase inhibitor clade C antithrombin member 1* (*SERPINC1*; Wang *et al.* 2006). Numerous additional genes found to be selectively expressed in liver have functions related to nutrient or metabolite transport, including *albumen* (*ALB*; Alpini *et al.* 1992), *group-specific component* (*vitamin D binding protein, GC*; Cooke *et al.* 1991), *haptoglobin* (*HP*; Oliviero *et al.* 1987), *transferrin* (*TF*; Idzerda *et al.* 1986), *ceruloplasmin* (*CP*; Aldred *et al.* 1987), *apolipoprotein C-IV* (*APOC4*; Zhang *et al.* 1996) and *alpha-1-microglobulin* (*AMBP*; Tyagi *et al.* 2002). In addition, members of the cytochrome P450 family of enzymes including *CYP2C9* and *CYP3A7*, which are involved in lipid oxidation pathways were also found to be selectively expressed in liver, consistent with observations for human liver (Hines 2007).

Tissue selective expression of genes expected based on their known functions or expression patterns in other species was observed for genes in brain and uterine endothelium. *Myelin basic protein* (*MBP*) is the major constituent of the myelin sheath for cells in the nervous system, and *MBP* transcript abundance is high in brain tissues (Kamholz *et al.* 1988). *Microtubule-associated protein 1B* (*MAP1B*) is a member of the microtubule-associated protein family thought to be involved with microtubule assembly which is essential for neurogenesis, and this gene is highly expressed in brain (Nunez & Fischer 1997). In addition, the *RIMS1* gene (*regulating synaptic membrane exocytosis 1*) expressed in neural tissues functions in regulating neurotransmitter

release (Lu *et al.* 2006). Genes revealed to be selectively expressed in the uterine endothelium included *serine peptidase inhibitor Kunitz type 2* (*placental bikunin, SPINT2*; Hettinger *et al.* 2001) and *S100 calcium binding protein A6* (*prolactin receptor-associated protein, S100A6*; Murphy *et al.* 1988).

In practice, microarray results are commonly validated using QPCR assays (Chuaqui *et al.* 2002; Morey *et al.* 2006). We confirmed through QPCR experiments that either total RNA or aRNA could be used for these tests; similar results have been reported in other species (Feldman *et al.* 2002; Li *et al.* 2002). As expected, genes often used as positive controls, *GAPDH* and *RPL32*, were expressed at high levels in almost all tissues. Tissue-specific gene expression was confirmed for *MBP;* in humans this gene is exclusively expressed in brain tissue and shows a similar expression pattern to the tissues tested here both by QPCR and by microarray analysis. Similarly, liver-specific expression was confirmed for *ALB* and *TF*, muscle-specific expression was confirmed for *GAPDH*, and uterine endothelium-specific expression was confirmed for *IGJ*.

A set of genes that exhibited significant differential expression in the microarray experiment was further evaluated by QPCR. Most of the significant differences ($P < 0.01$) detected in the microarray experiment were replicated through QPCR analysis. Genes that failed to replicate between the array and QPCR results (i.e. *INDO* and *HSPA8*) exhibited relatively low signal intensities on the microarray. In total, seven of the 10 differentially expressed genes detected by the microarray ($P < 0.01$) were confirmed by QPCR with significant correlations in expression patterns between the microarray and QPCR ($P < 0.01$), and another gene exhibited a similar expression pattern for QPCR and microarray, with the correlation in the correct direction although not significant ($P > 0.1$).

In summary, we tested the Pigoligoarray and verified its specificity and validity using hybridization intensity diagnostics and assessment of tissue enhanced gene expression. Simple descriptive diagnostic analyses of PM/MM probe sets introduced in this paper are useful to detect non-specific hybridization. Using comparative transcriptional profiling, we found that the microarray data correlate to QPCR data for most genes detected to be differentially expressed using the microarray platform. Moreover, comparison to human ortholog expression confirmed the value of this array for experiments of both agricultural importance and for using pigs as a biomedical model for human disease.

Acknowledgements

References

Aberle E.D., Forrest J.C., E G.D. & W M.E. (2001) Structure and composition of animal tissues. In: *Principles of Meat Science* (Ed. by E.D. Aberle), pp. 9–20. Kendall/Hunt, Dubuque, Iowa.

Aldred A.R., Grimes A., Schreiber G. & Mercer J.F. (1987) Rat ceruloplasmin. Molecular cloning and gene expression in liver, choroid plexus, yolk sac, placenta, and testis. *Journal of Biological Chemistry* **262**, 2875–8.

Alpini G., Aragona E., Dabeva M., Salvi R., Shafritz D.A. & Tavoloni N. (1992) Distribution of albumin and alpha-fetoprotein mRNAs in normal, hyperplastic, and preneoplastic rat liver. *American Journal of Pathology* **141**, 623–32.

Babu G.J., Bhupathy P., Carnes C.A., Billman G.E. & Periasamy M. (2007) Differential expression of sarcolipin protein during muscle development and cardiac pathophysiology. *Journal of Molecular and Cellular Cardiology* **43**, 215–22.

Chuaqui R.F., Bonner R.F., Best C.J. *et al.* (2002) Post-analysis follow-up and validation of microarray experiments. *Nature Genetics* **32**(Suppl.), 509–14.

Cooke N.E., McLeod J.F., Wang X.K. & Ray K. (1991) Vitamin D binding protein: genomic structure, functional domains, and mRNA expression in tissues. *Journal of Steroid Biochemistry and Molecular Biology* **40**, 787–93.

Cui X. & Churchill G.A. (2003) How many mice and how many arrays? Replication in mouse cDNA microarray experiments. In: *Methods of Microarray Data Analysis III* (Ed. by K.F. Jonhson & S.M. Lin), pp. 139–54. Springer, New York.

Currier A.R., Sabla G., Locaputo S., Melin-Aldana H., Degen J.L. & Bezerra J.A. (2003) Plasminogen directs the pleiotropic effects of uPA in liver injury and repair. *American Journal of Physiology. Gastrointestinal and Liver Physiology* **284**, G508–15.

Dawson H.D., Beshah E., Nishi S. *et al.* (2005) Localized multigene expression patterns support an evolving Th1/Th2-like paradigm in response to infections with *Toxoplasma gondii* and *Ascaris suum*. *Infection and Immunity* **73**, 1116–28.

Duan H.O. & Simpson-Haidaris P.J. (2006) Cell type-specific differential induction of the human gamma-fibrinogen promoter by interleukin-6. *Journal of Biological Chemistry* **281**, 12451–7.

Feldman A.L., Costouros N.G., Wang E., Qian M., Marincola F.M., Alexander H.R. & Libutti S.K. (2002) Advantages of mRNA amplification for microarray analysis. *BioTechniques* **33**, 906–12.

Gadbury G.L., Page G.P., Edwards J., Kayo T., Prolla T.A., Weindruch R., Permana P.A., Mountz J.D. & Allison D.B. (2004) Power and sample size estimation in high dimensional biology. *Statistical Methods in Medical Research* **13**, 325–38.

Hettinger A.M., Allen M.R., Zhang B.R., Goad D.W., Malayer J.R. & Geisert R.D. (2001) Presence of the acute phase protein, bikunin,

in the endometrium of gilts during estrous cycle and early pregnancy. *Biology of Reproduction* **65**, 507–13.

Hines R.N. (2007) Ontogeny of human hepatic cytochromes P450. *Journal of Biochemical and Molecular Toxicology* **21**, 169–75.

Hu Z.L., Bao J. & Reecy J.M. (2008) CateGOrizer: a web-based program to batch analyze gene ontology classification categories. *Online Journal of Bioinformatics* **9**, 108–12.

Idzerda R.L., Huebers H., Finch C.A. & Mcknight G.S. (1986) Rat *transferrin* gene-expression – tissue-specific regulation by iron-deficiency. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 3723–7.

Kamholz J., Toffenetti J. & Lazzarini R.A. (1988) Organization and expression of the human *myelin basic-protein* gene. *Journal of Neuroscience Research* **21**, 62–70.

Li Y., Li Y.L., Tang R. *et al.* (2002) Discovery and analysis of hepatocellular carcinoma genes using cDNA microarrays. *Journal of Cancer Research and Clinical Oncology* **128**, 369–79.

Lu H.C., Butts D.A., Kaeser P.S., She W.C., Janz R. & Crair M.C. (2006) Role of efficient neurotransmitter release in barrel map development. *Journal of Neuroscience* **26**, 2692–703.

Lunney J.K. (2007) Advances in swine biomedical model genomics. *International Journal of Biological Sciences* **3**, 179–84.

Morey J.S., Ryan J.C. & Van Dolah F.M. (2006) Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. *Biological Procedures Online*, **8**, 175–93.

Morisaki T. & Holmes E.W. (1993) Functionally distinct elements are required for expression of the *AMPD1* gene in myocytes. *Molecular and Cellular Biology* **13**, 5854–60.

Mukai T., Joh K., Arai Y., Yatsuki H. & Hori K. (1986) Tissue-specific expression of rat *aldolase-a* messenger-RNAas - 3 molecular-species differing only in the 5′-terminal sequences. *Journal of Biological Chemistry* **261**, 3347–54.

Murphy L.C., Murphy L.J., Tsuyuki D., Duckworth M.L. & Shiu R.P.C. (1988) Cloning and characterization of a cDNA-encoding a highly conserved, putative calcium-binding protein, identified by an anti-prolactin receptor antiserum. *Journal of Biological Chemistry* **263**, 2397–401.

Nunez J. & Fischer I. (1997) Microtubule-associated proteins (MAPs) in the peripheral nervous system during development and regeneration. *Journal of Molecular Neuroscience* **8**, 207–22.

Oliviero S., Morrone G. & Cortese R. (1987) The human *haptoglobin* gene – transcriptional regulation during development and acute phase induction. *EMBO Journal* **6**, 1905–12.

Petry D.B., Lunney J., Boyd P., Kuhar D., Blankenship E. & Johnson R.K. (2007) Differential immunity in pigs with high and low responses to porcine reproductive and respiratory syndrome virus infection. *Journal of Animal Science* **85**, 2075–92.

Ragusa M.A., Costa S., Cefalu A.B., Noto D., Fayer F., Travali S., Averna M.R. & Gianguzza F. (2006) RT-PCR and *in situ* hybridization analysis of apolipoprotein H expression in rat normal tissues. *International Journal of Molecular Medicine* **18**, 449–55.

Rosa G.J.M., Steibel J.P. & Tempelman R.J. (2005) Reassessing design and analysis of two-colour microarray experiments using mixed effects models. *Comparative and Functional Genomics* **6**, 123–31.

Shen X.H., Franzini-Armstrong C., Lopez J.R. *et al.* (2007) Triadins modulate intracellular Ca2+ homeostasis but are not essential for excitation-contraction coupling in skeletal muscle. *Journal of Biological Chemistry* **282**, 37864–74.

Storey J.D. & Tibshirani R. (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9440–5.

Su A.I., Wiltshire T., Batalov S. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 6062–7.

Tsai S., Cassady J.P., Freking B.A., Nonneman D.J., Rohrer G.A. & Piedrahita J.A. (2006) Annotation of the Affymetrix(1) porcine genome microarray. *Animal Genetics* **37**, 423–4.

Tuggle C.K., Wang Y.F. & Couture O. (2007) Advances in swine transcriptomics. *International Journal of Biological Sciences* **3**, 132–52.

Tweedie S. & Edwards Y. (1989) Mouse carbonic-anhydrase. 3. Nucleotide-sequence and expression studies. *Biochemical Genetics* **27**, 17–30.

Tyagi S., Salier J.P. & Lal S.K. (2002) The liver-specific human alpha(1)-microglobulin/bikunin precursor (AMBP) is capable of self-association. *Archives of Biochemistry and Biophysics* **399**, 66–72.

\Uenishi H., Eguchi T., Suzuki K., Sawazaki T., Toki D., Shinkai H., Okumura N., Hamasima N. & Awata T. (2004) PEDE (Pig EST Data Explorer): construction of a database for ESTs derived from porcine full-length cDNA libraries. *Nucleic Acids Research* **32**, D484–8.

Uenishi H., Eguchi-Ogawa T., Shinkai H., Okumura N., Suzuki K., Toki D., Hamasima N. & Awata T. (2007) PEDE (Pig EST Data Explorer) has been expanded into Pig Expression Data Explorer, including 10 147 porcine full-length cDNA sequences. *Nucleic Acids Research* **35**, D650–3.

Wang A.G., Yoon S.Y., Oh J.H. *et al.* (2006) Identification of intrahepatic cholangiocarcinoma related genes by comparison with normal liver tissues using expressed sequence tags. *Biochemical and Biophysical Research Communications* **345**, 1022–32.

Wang Y.F., Qu L., Uthe J.J., Bearson S.M.D., Kuhar D., Lunney J.K., Couture O.P., Nettleton D., Dekkers J.C.M. & Tuggle C.K. (2007) Global transcriptional response of porcine mesenteric lymph nodes to *Salmonella enterica* serovar Typhimurium. *Genomics* **90**, 72–84.

Yang Y.H., Dudoit S., Luu P., Lin D.M., Peng V., Ngai J. & Speed T.P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research* **30**, e15.

Zhang L.H., Kotite L. & Havel R.J. (1996) Identification, characterization, cloning, and expression of apolipoprotein C-IV, a novel sialoglycoprotein of rabbit plasma lipoproteins. *Journal of Biological Chemistry* **271**, 1776–83.

Zhao S.H., Recknor J., Lunney J.K., Nettleton D., Kuhar D., Orley S. & Tuggle C.K. (2005) Validation of a first-generation long-oligonucleotide microarray for transcriptional profiling in the pig. *Genomics* **86**, 618–25.

Supporting information

Additional supporting information may be found in the online version of this article.

Figure S1 Loop design used in the microarray experiment. The arrows represent arrays, the tails point at the sample labelled with the green dye, while the head points at the sample labelled with the red dye.

Figure S2 Average intensity across negative probes in five experiments.

Figure S3 A. Histogram of *P*-values. B. The expected proportion of false positives as a function of the number of significant tests. C. Number of significant tests for different *q*-value thresholds.

Table S1 A. Description of experimental datasets used to assess the intensity of negative probes. B. Average intensities (A-values) of 60 oligonucleotide probes in the five independent experiments.

Table S2 Comparison of assigned Gene Ontology (GO) terms for commercially available swine arrays.

Table S3 (**a**) Top 15 differentially over-expressed genes in longissimus muscle. (**b**) Top 15 differentially over-expressed genes in liver.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.